Validating the Site Productivity Layer for British Columbia with Equivalence Testing

2015

Validating the Site Productivity Layer for British Columbia with Equivalence Testing

Gordon Nigh and Rene de Jong

The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the Government of British Columbia of any product or service to the exclusion of any others that may also be suitable. Contents of this report are presented for discussion purposes only. Funding assistance does not imply endorsement of any statements or information contained herein by the Government of British Columbia. Uniform Resource Locators (URLs), addresses, and contact information contained in this document are current at the time of printing unless otherwise noted.

Print edition: ISBN 978-0-7726-6867-7
Electronic/PDF edition: ISBN 978-0-7726-6868-4

Citation
Nigh, G.D. and R. de Jong. 2015. Validating the site productivity layer for British Columbia with equivalence testing. Prov. B.C., Victoria, B.C. Tech. Rep. 085.
www.for.gov.bc.ca/hfd/pubs/Docs/Tr/Tr085.htm

Authors' affiliation
Gordon Nigh and Rene de Jong
B.C. Ministry of Forests, Lands and Natural Resource Operations
Forest Analysis and Inventory Branch
PO Box 9512, Stn Prov Govt
Victoria, BC v8w 9c2

Copies of this report may be obtained from:
Crown Publications, Queen's Printer
PO Box 9452 Stn Prov Govt
Victoria, BC v8w 9v7
1-800-663-6105
www.crownpub.bc.ca

For information on other publications in this series, visit
www.for.gov.bc.ca/scripts/hfd/pubs/hfdcatalog/index.asp

© 2015 Province of British Columbia
When using information from this report, please cite fully and correctly.

ABSTRACT

The site productivity layer is a database of site indexes for commercial species, which can be linked to a particular area (a site) in the province. It was created in response to a need for improved accessibility to site index information. The site index estimates in the layer were obtained from either the Site Index – Biogeoclimatic Ecosystem Classification (SIBEC) model coupled with an ecological map, or from a biophysical model. This document reports on the validation of the site productivity layer. The validation data consist of site indexes for various species from inventory plots (e.g., monitoring plots, Growth Natural Permanent Sample Plots, Vegetation Resources Inventory plots) and research and silviculture plots. These site indexes were paired with site index estimates from the site productivity layer for the same location and species. Equivalence testing, which tests for overall bias and 1:1 correspondence between the predicted and observed site indexes, was used to perform the validation. The equivalence test was applied to the complete layer and to strata created from the layer. These strata were defined by species, species (including all species combined) and model type (biophysical or SIBEC), and species and biogeoclimatic zone. As a whole, the layer was validated. However, many of the strata were not validated by the equivalence testing. Slightly more than half of the strata passed the equivalence test for unbiasedness but few strata were validated for 1:1 correspondence between the predicted and observed site indexes.

ACKNOWLEDGEMENTS

We thank Graham Hawkins, Deb MacKillop, Shirley Mah, Peter Ott, Paul Rehsler, and Michael Ryan (all with the British Columbia Ministry of Forests, Lands and Natural Resource Operations), Michael Peterman (from CloverPoint), and Jeff Kruys (formerly with CloverPoint) for valuable review comments.

CONTENTS

Abstract iii
Acknowledgements iii
Introduction 1
Data 3
Methods 5
Results 7
Discussion 22
Conclusion 23
Literature Cited 24

TABLES

1 List of species in the site productivity layer that were validated, along with their biogeoclimatic zones and number of observations 2
2 Results of the equivalence testing 8

FIGURES

1 Plots of ground-based site index versus layer site index by species 12

The site productivity layer project was initiated in 2011 as a Land Based Investment Strategy project. The objective was to improve accessibility to site productivity information. This was done by creating a database of site indexes for commercial species, which can be linked to a particular area (a site) in the province, where site index is the height of a site tree at breast height age 50 and is a measure of site productivity. The database of site indexes is referred to as the site productivity layer ("the layer") since it is in a Geographic Information System (GIS) format and can be accessed as a GIS layer. The first operational release of the layer occurred in 2013 with version 3.0. The second and latest (at the time of writing) release was in 2014 with version 3.1 (British Columbia Ministry of Forests, Lands and Natural Resource Operations 2014a). The intended use of the layer is for strategic, as opposed to operational, purposes (CloverPoint 2014).

The layer conceptually divides British Columbia into 1-ha cells using the same grid cells as Hectares BC (Hectares BC 2014). A site index for each species potentially growing in the grid cell is estimated and assigned to the cell based on the ecological, biophysical, and/or climatic conditions at the centroid of the cell. Each grid cell can have a site index for up to 22 species. Usually cells have much fewer than 22 site indexes because fewer species grow in that area of the province and/or because site index estimates are lacking for some species. The species in the site productivity layer that are validated are listed in Table 1.

There are two sources of site index estimates in the layer. The first is the SIBEC model (British Columbia Ministry of Forests and Range 2011) coupled with a predictive ecosystem map (PEM) or a terrestrial ecosystem map (TEM) (British Columbia Ministry of Environment 2014). The PEM or TEM is used to predict the site series (i.e., the ecosystem type) at the centroid of the grid cell. The SIBEC model is then queried to determine all the species present and their respective site indexes for that site series. Often the site series occur in complexes. These are areas that are heterogeneous with respect to site series but where the individual site series in the complex are too small to be mapped separately. In cases where the centroid of the cell is in a PEM or TEM complex, the site index in the layer is a weighted average of the site indexes for the site series in the complex, where the weighting is based on the area of the site series. A complex can contain up to three site series.

The other source of site index estimates is the biophysical model (Nigh 2012), which is a regression equation that predicts site index from biogeoclimatic zone (Meidinger and Pojar 1991; British Columbia Ministry of Forests and Range 2014) and biophysical variables, such as slope, aspect, elevation, and climate variables. A model is available for those species listed in Table 1, each of which has a unique set of predictor variables. The climate predictor variables are obtained from ClimateWNA (Wang et al. 2011), which is downloadable from the Centre for Forest Conservation Genetics (2014).

This report describes the results of a validation of the site productivity layer. Rykiel (1996) explores the concept of validation in some depth; his ideas are borrowed for defining validation. He defines validation as "deciding if the model is acceptable for its intended use, i.e., whether the model mimics the real world well enough for its stated purpose" or "a demonstration that a

TABLE 1 *List of species in the site productivity layer that were validated, along with their biogeoclimatic zones (Meidinger and Pojar 1991) and number of observations*

Species code	Common name	Scientific name	Biogeoclimatic zones[a]	No. of observations
At	Trembling aspen	*Populus tremuloides*	BWBS	266
			ICH	35
			SBS	215
			Other	21
Ba	Amabilis fir	*Abies amabilis*	CWH	203
			Other	4
Bl	Subalpine fir	*Abies lasiocarpa*	ESSF	428
			ICH	76
			MS	68
			SBS	280
			Other	1
Cw	Western redcedar	*Thuja plicata*	CWH	245
			Other (coast)	4
			ICH	175
			Other (interior)	17
Ep	Paper birch	*Betula papyrifera*	ICH	44
			SBS	49
			Other	8
Fd	Douglas-fir	*Pseudotsuga menziesii*	CDF	97
			CWH	884
			Other (coast)	3
			ICH	404
			IDF	908
			MS	79
			SBS	143
			Other (interior)	21
Hw	Western hemlock	*Tsuga heterophylla*	CWH	1289
			Other (coast)	6
			ICH	294
			Other (interior)	2
Lw	Western larch	*Larix occidentalis*	ICH	94
			IDF	60
			MS	35
Pl	Lodgepole pine	*Pinus contorta*	BWBS	135
			ESSF	253
			ICH	278
			IDF	317
			MS	428
			SBPS	288
			SBS	731
Py	Ponderosa pine	*Pinus ponderosa*	PP	28
			Other	19
Sb	Black spruce	*Picea mariana*	BWBS	70
			Other	11
Se	Engelmann spruce	*Picea engelmannii*	ESSF	157
Ss	Sitka spruce	*Picea sitchensis*	CWH	235
Sw	White spruce	*Picea glauca*	BWBS	164
Sx	Interior spruce	*Picea glauca* × *engelmanii*	ICH	127
			IDF	28
			MS	46
			SBS	594
			Other	16

a BWBS – Boreal White and Black Spruce; CDF – Coastal Douglas-Fir; CWH – Coastal Western Hemlock; ESSF – Engelmann Spruce – Subalpine Fir; ICH – Interior Cedar – Hemlock; IDF – Interior Douglas-Fir; MS – Montane Spruce; PP – Ponderosa Pine; SBPS – Sub-Boreal Pine – Spruce; SBS – Sub-Boreal Spruce; Other – zones that had fewer than 25 observations were grouped into this class.

model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (Rykiel 1996). These definitions are important because they implicitly define three factors that were considered when validating the layer (the model): (1) the intended use of the model, (2) criteria that must be met in order for the model to be declared fit to use (e.g., how much uncertainty can be tolerated for a given application), and (3) the context (e.g., assumptions, restrictions on use) in which the model is intended to operate (Rykiel 1996). The validation was done using equivalence testing (Barker et al. 2002; Blair and Cole 2002; Wellek 2010) because this procedure meets the above factors for validating a model. In particular, the second factor above is explicitly stated with equivalence testing.

DATA

The data for the validation of the layer were obtained from several sources. The following programs provided sample plot data: Change Monitoring Inventory (British Columbia Ministry of Forests, Lands and Natural Resource Operations 2014b), Growth Natural Permanent Sample Plots (British Columbia Ministry of Forests, Lands and Natural Resource Operations 2014c), Intensive Forestry Program (British Columbia Ministry of Forests 1994), research plots, Silviculturally Treated Program (British Columbia Ministry of Forests 1994), Vegetation Resource Inventory (British Columbia Ministry of Forests, Lands and Natural Resource Operations 2014d), Young Stand Monitoring (British Columbia Ministry of Forests, Lands and Natural Resource Operations 2014b), and other monitoring programs (British Columbia Ministry of Forests, Lands and Natural Resource Operations 2014b). The research plots were from various research projects; only control plots from these projects were analyzed.

Site index was estimated from height and breast height age data for site tree(s) selected from each plot. In some programs, site tree selection was done at the time of plot measurement. In other programs, the site tree was selected in an office exercise using the suitability criteria for site tree selection in the SIBEC sampling standards (British Columbia Ministry of Forests and Range 2011). In brief, a site tree is the largest-diameter tree of the target species in a 0.01-ha plot, and must be undamaged and free of any non-site influences. Site trees that were older than 120 years at breast height age or younger than 10 years (or 20 years if a growth intercept model was not available for that species) were removed from the data set. The standards were designed to ensure that the height growth of the site tree reflects the potential productivity of the site. If the plot was 0.01 ha, the site index was calculated from the breast height age and height of the single site tree. For plots that were larger than 0.01 ha, the site trees were selected and site index was calculated in the following three ways:

1. For plots that were sectored, the largest-diameter tree in each sector was selected as the site tree, provided the tree met the suitability criteria. Sectoring is the process of dividing the plot into approximate 0.01-ha slices to

ease tree tagging and measuring. The site index was calculated from breast height age and height for each site tree and then averaged for the plot.

2. The *n* largest-diameter trees were selected as site trees, provided they met the suitability criteria, where *n* = plot size/0.01. Plot sectoring, if done, was ignored. The site index for each site tree was calculated from breast height age and height and then averaged to give a site index for the plot.
3. The *n* largest-diameter trees were selected as site trees (ignoring sectoring), provided they met the suitability criteria, where *n* = plot size/0.01. The breast height ages and heights of the site trees were averaged, and the plot site index was then calculated from the average height and age.

When site index was calculated more than one way, the order of preference for choosing which site index to analyze was the same as outlined in the numbered list above. For plots with multiple measurements, the site index was calculated from the measurement closest to the index age of 50 years at breast height. In all situations, the site index was calculated using the recommended site index curves in the SiteTools software (British Columbia Ministry of Forests and Range 2001) with version 1.45 of the Sindex.dll.

The spruce species were coded to be consistent with the SIBEC model and the biophysical model. All spruce in the CWH biogeoclimatic zone were coded as Sitka spruce except for spruce coded as Sx in the CWHws subzone, which retained their coding of Sx. All black spruce not in the CWH biogeoclimatic zone remained coded as black spruce. Of the remaining spruce species, trees in the ESSF biogeoclimatic zone were coded as Engelmann spruce, trees in the BWBS biogeoclimatic zone were coded as white spruce, and other spruce trees were coded as interior spruce.

The biogeoclimatic zone was obtained from the January 2014 provincial biogeoclimatic maps (British Columbia Ministry of Forests and Range 2014).

The site index from the layer was obtained by first identifying the cell in the layer in which the validation plot was established. The layer was then queried and the site index was extracted for the species for which a site index was calculated from the validation plot. The model source (SIBEC or biophysical) for the site index was also extracted. In a few cases, the grid cell in the layer was missing data. In these cases, the grid size was increased to 4 ha (i.e., four cells) and the site index was averaged from these cells. This led to some site index estimates being obtained from both the SIBEC and biophysical models. These site indexes were given a model source of "Other." Observations for unique combinations of species and biogeoclimatic zones with fewer than 25 observations were grouped into a biogeoclimatic zone designated as "Other."

The analysis-ready data set contained a ground-based site index (considered to be the true site index), the estimated site index from the layer, species, biogeoclimatic zone, and model source for the estimated site index. Table 1 lists the species, biogeoclimatic zones, and number of observations in the validation data set. Only species with more than 25 observations were used in the validation.

A common strategy for validation is to fit a regression line between the model predictions and the observations and test whether the slope and intercept of the line differ from 1 and 0, respectively. There are two problems with this approach (Parkhurst 2001; Barker et al. 2002; Robinson and Froese 2004; Robinson et al. 2005). First, a failure to reject the null hypothesis does not mean that the null hypothesis must be accepted; it could also be that there is not enough evidence (data) to reject the null hypothesis. Second, the standard errors of the estimates for the slope and intercept parameters decrease as the sample size increases; consequently, the probability of rejecting the null hypothesis increases as the sample size increases even though the difference between the predictions and observations may be inconsequential.

Equivalence testing overcomes the above-mentioned problems (Parkhurst 2001; Blair and Cole 2002; Robinson and Froese 2004). The following example outlines the basic principles of equivalence testing. Suppose the responses from two treatments are to be examined for equivalence. The null hypothesis is one of dissimilarity; that is, the null hypothesis to be tested is that the difference between the two mean responses is at least some magnitude apart (Blair and Cole 2002). An equivalence interval is specified such that if the difference between the two means is in the equivalence interval, then the null hypothesis is rejected and the means are declared to be equivalent (similarity is established) (defining an equivalence interval is a somewhat subjective process, but Wellek 2010 provides some guidelines). Two one-sided tests (TOSTs) are conducted to establish equivalence. One tests if the difference in means is below the lower bound of the equivalence interval; the other tests if the difference is above the upper bound. Both tests must be rejected at the pre-specified significance level in order to declare equivalence. Therefore, if α is the chosen significance level (typically $\alpha = 0.05$), then the tests must be conducted at a $2 \times \alpha$ significance level (Barker et al. 2001; Blair and Cole 2002).

In the context of validating a model against an independent data set, Robinson et al. (2005) establish a procedure to test for equivalence. It is assumed that a model validation data set has been collected and that there are predictions from the model (the layer, in this case) that correspond with observations in the validation data set. The two null hypotheses being tested are that the means of the predictions and observations are different and that the slope of the linear regression line between the observations and predictions is different from 1. The first hypothesis is a test for bias; the second is a test for 1:1 correspondence between the predictions and observations. The equivalence test procedure is briefly described here:

1. Subtract the mean of the predicted values from each predicted value.
2. Determine an equivalence interval for the shifted intercept and the slope.
3. Fit a linear regression with the model predictions as the independent variable and the observations as the dependent variable.
4. Use a TOST to test if the intercept is equivalent to the mean of the predictions (the fitted intercept is now the mean of the observations, and if the model is unbiased, the intercept will be equivalent to the mean of the predictions).
5. Use a TOST to test if the slope is equivalent to 1.

Since two comparisons are made (one for the slope and one for the intercept), α must be corrected to account for multiple comparisons (Robinson et al. 2005). If the null hypotheses for the test for both the slope and the intercept are rejected, then the model is deemed to have successfully passed the validation test.

The confidence intervals used in the TOST are based on the usual assumptions of independence, constant variance, and normality. Robinson et al. (2005) suggest using a non-parametric bootstrap procedure to construct confidence intervals that do not depend on these assumptions to safeguard against coming to the wrong conclusion if these assumptions are not met (knowingly or unknowingly).

The "equivalence" package in the R software (R Core Team 2013) was used to validate the layer. The function equiv.boot was applied to perform the equivalence test on the model predictions and observations with a bootstrapped confidence interval. The size of the test was set at $\alpha = 0.05$, and 10,000 replicates were specified for the bootstrap. The equivalence interval was set at ±10% for the intercept and the slope. While there is no definitive reason why 10% was used, it was thought that this creates an equivalence interval that is not too lax but is also not so stringent that the layer is validated only when the predictions are virtually perfect. With the chosen equivalence intervals, the null hypotheses were rejected if the confidence interval for the intercept was within ±10% of the mean of the predictions and the confidence interval for the slope was between 0.9 and 1.1. Enough information is presented in the Results section to calculate other equivalence intervals and perform equivalence testing with those intervals.

In cases where the null hypotheses were not rejected, the equivalence interval was increased in steps of 5% until the null hypothesis for the intercept and/or the slope was rejected. This gives an approximate smallest equivalence interval in which the hypothesis of dissimilarity is rejected.

The layer was stratified and an equivalence test was performed on each stratum. The strata were as follows:

1. The layer was validated as one unit to give an overall picture of the performance of the layer.
2. Each species was validated individually (see Table 1 for a list of species that were validated).
3. The layer was validated by species and model type (SIBEC or biophysical). Note that observations from both models (see Data section) were not included in this analysis.
4. Validation was done by species and biogeoclimatic zone. Observations in the "Other" category of biogeoclimatic zone were not included in this analysis.

Graphs of the observed site indexes versus the predicted site indexes were constructed to show the relationships between these two variables. One graph was produced for all observations, and a series of graphs was produced by species. Different icons were used to represent estimates from the different models, and colours were used to show the biogeoclimatic zones in the species-specific graphs. The 1:1 line (the line representing perfect correspondence between observations and predictions) was also drawn on the graphs.

The results of the equivalence tests are presented according to the different strata for which the analysis was done (Table 2). When examining the table, recall that:

- The null hypothesis for equivalence testing is that the validation data and the layer are dissimilar; a rejection of the null hypothesis leads to the conclusion that the data and the layer are similar.
- The equivalence test for the intercept is a test for unbiasedness.
- The equivalence test for the slope is for point-to-point agreement between the layer and the validation data. That is, it tests whether layer predictions with high site indexes correspond with plots with high site indexes, and vice versa.
- In order to pass the validation, the null hypothesis for both tests should be rejected, although it may be the case for some applications that the test for unbiasedness is more important than the test for point-to-point correspondence.

The data in Table 2 are interpreted as follows. The estimates of the slope and the intercept are presented along with their 95% bootstrapped confidence intervals. The equivalence interval for the intercept is the mean of the predicted site indexes from the layer ± 10%; these three values are also in the table. The equivalence interval for the slope is 1.0 ± 10% (0.9–1.1); these values are common across all tests and hence are not reported in the table. The null hypothesis for equivalence testing is rejected if the 95% confidence interval for the intercept and the slope is wholly contained within the equivalence interval. If the null hypothesis for the intercept is rejected, then the alternative hypothesis of similarity is accepted and the layer is deemed to be unbiased. Similarly for the slope, if the null hypothesis is rejected, we conclude that there is good correspondence between the layer and the actual site indexes. Also presented in the table for the tests that are not rejected with a 10% equivalence interval is the smallest equivalence interval (as a percent of the mean to the nearest 5%) for which the null hypothesis is rejected.

The graphs in Figure 1 show the relationship between the measured site index from the inventory plots and the predicted site index from the layer.

TABLE 2 *Results of the equivalence testing*

		Intercept							Slope					
		95% C.I.		Equivalence interval				Smallest equivalence		95% C.I.			Smallest equivalence	No. of
Stratum	Estimate	Lower	Upper	Mean	Lower	Upper	Reject?	interval (±%)	Estimate	Lower	Upper	Reject?	interval (±%)	obs.
By species[a]														
All	20.44	20.35	20.53	21.33	19.20	23.47	Y		0.982	0.963	1.002	Y		10 383
Coast														
Ba	23.86	23.16	24.53	25.25	22.73	27.78	Y		0.800	0.588	1.035	N	45	207
Cw	20.91	20.28	21.53	21.40	19.26	23.54	Y		0.546	0.367	0.732	N	65	249
Fd	29.76	29.38	30.13	32.14	28.92	35.35	Y		0.601	0.490	0.713	N	55	984
Hw	27.97	27.65	28.30	25.49	22.94	28.04	N	15	0.581	0.463	0.700	N	55	1295
Ss	32.04	31.34	32.72	28.69	25.82	31.56	N	15	0.384	0.045	0.756	N	100	235
Interior														
At	18.87	18.58	19.17	18.42	16.58	20.26	Y		0.360	0.200	0.515	N	80	537
Bl	14.11	13.83	14.39	16.86	15.17	18.55	N	20	0.523	0.418	0.631	N	60	853
Cw	17.57	16.95	18.18	17.35	15.61	19.08	Y		0.478	0.038	0.906	N	100	192
Ep	18.73	18.21	19.27	19.30	17.37	21.23	Y		0.471	-0.195	1.101	N	120	101
Fd	17.96	17.77	18.16	19.87	17.89	21.86	N	15	0.814	0.746	0.882	N	30	1555
Hw	17.97	17.51	18.42	19.05	17.14	20.95	Y		0.580	0.253	0.885	N	80	296
Lw	20.97	20.50	21.43	21.92	19.72	24.11	Y		0.549	0.357	0.756	N	65	189
Pl	17.15	17.02	17.28	18.80	16.92	20.68	Y		0.957	0.905	1.008	Y		2430
Py	14.18	13.20	15.10	16.00	14.40	17.60	N	20	1.168	0.159	3.176	N	220	47
Sb	10.71	10.07	11.38	10.34	9.30	11.37	N	15	0.321	-0.359	0.991	N	135	81
Se	16.07	15.39	16.75	16.14	14.53	17.75	Y		0.240	-0.003	0.471	N	105	157
Sw	18.62	18.01	19.23	19.08	17.17	20.99	Y		0.364	0.068	0.673	N	95	164
Sx	19.28	19.03	19.53	20.25	18.22	22.27	Y		0.605	0.460	0.753	N	55	811
By species and model														
All, B	22.03	21.85	22.21	23.23	20.91	25.55	Y		0.961	0.927	0.994	Y		3064
All, S	19.73	19.62	19.83	20.49	18.44	22.54	Y		1.002	0.976	1.027	Y		7183
Coast														
Ba, B	23.07	22.09	23.98	24.94	22.44	27.43	N	15	1.452	1.168	1.720	N	75	104
Ba, S	24.66	23.77	25.51	25.58	23.02	28.13	Y		0.525	0.310	0.763	N	70	103
Cw, B	19.75	18.83	20.65	20.08	18.07	22.09	Y		0.526	0.188	0.881	N	85	113
Cw, S	22.00	21.16	22.84	22.54	20.29	24.80	Y		0.441	0.220	0.686	N	80	131
Fd, B	29.39	28.88	29.90	32.89	29.60	36.18	N	15	0.967	0.689	1.226	N	35	523

Stratum	Intercept								Slope						No. of obs.
	Estimate	95% C.I.		Equivalence interval			Reject?	Smallest equivalence interval (±%)	Estimate	95% C.I.		Reject?	Smallest equivalence interval (±%)		
		Lower	Upper	Mean	Lower	Upper				Lower	Upper				
Fd, S	30.14	29.60	30.68	31.23	28.11	34.36	Y		0.610	0.492	0.728	N	55		448
Hw, B	26.76	26.26	27.28	25.46	22.92	28.01	Y		1.426	1.176	1.674	N	70		518
Hw, S	28.70	28.28	29.11	25.40	22.86	27.94	N	15	0.402	0.271	0.534	N	75		741
Ss, B	30.72	29.86	31.60	28.79	25.91	31.67	Y		1.180	0.737	1.612	N	65		124
Ss, S	33.57	32.46	34.63	28.24	25.42	31.07	N	25	0.090	-0.570	0.730	N	160		98
Interior															
At, B	18.42	17.89	18.94	18.89	17.00	20.78	Y		0.462	0.230	0.680	N	80		176
At, S	19.08	18.73	19.45	18.17	16.36	19.99	Y		0.380	0.169	0.594	N	85		354
Bl, B	13.30	12.73	13.86	16.35	14.71	17.98	N	25	1.163	0.907	1.416	N	45		204
Bl, S	14.39	14.06	14.71	17.01	15.31	18.71	N	20	0.397	0.284	0.510	N	75		639
Cw, B	18.09	16.62	19.48	17.23	15.51	18.95	N	15	-0.416	-2.640	1.532	N	370		35
Cw, S	17.57	16.90	18.19	17.38	15.64	19.11	Y		0.533	0.010	0.961	N	95		155
Ep, S	18.84	18.23	19.47	19.29	17.37	21.22	Y		0.627	-0.213	1.406	N	120		77
Fd, B	18.41	18.01	18.82	20.29	18.26	22.32	N	15	1.148	0.978	1.317	N	35		337
Fd, S	17.85	17.64	18.08	19.77	17.79	21.74	N	15	0.738	0.667	0.809	N	35		1213
Hw, B	17.81	17.09	18.49	19.11	17.20	21.02	N	15	0.414	-0.062	0.866	N	110		125
Hw, S	18.07	17.44	18.67	19.02	17.12	20.92	Y		0.684	0.210	1.150	N	80		168
Lw, B	20.15	19.59	20.71	22.01	19.81	24.22	N	15	1.121	0.754	1.482	N	50		75
Lw, S	21.50	20.86	22.11	21.85	19.67	24.04	Y		0.404	0.182	0.624	N	85		114
Pl, B	18.23	17.95	18.52	19.90	17.91	21.89	Y		1.122	1.000	1.243	N	25		462
Pl, S	16.89	16.74	17.04	18.53	16.67	20.38	Y		0.925	0.867	0.982	N	15		1948
Py, B	14.67	13.64	15.99	16.11	14.50	17.72	N	20	1.920	0.801	5.229	N	425		25
Sb, S	10.87	10.14	11.65	10.45	9.40	11.49	N	15	0.414	-0.416	1.247	N	145		55
Se, B	16.16	14.62	17.69	16.51	14.86	18.16	N	15	0.439	-0.154	1.006	N	115		41
Se, S	16.04	15.32	16.77	16.01	14.41	17.61	Y		0.193	-0.068	0.454	N	110		116
Sw, B	18.66	17.21	19.87	19.73	17.75	21.70	N	15	-0.068	-0.557	0.322	N	155		26
Sw, S	18.66	17.96	19.33	18.96	17.06	20.85	Y		0.455	0.097	0.826	N	90		137
Sx, B	20.26	19.76	20.76	20.83	18.74	22.91	Y		0.724	0.438	0.990	N	60		135
Sx, S	19.16	18.87	19.43	20.13	18.12	22.15	Y		0.547	0.388	0.712	N	65		664

TABLE 2 *Continued*

Stratum	Estimate	Intercept 95% C.I. Lower	Intercept 95% C.I. Upper	Equivalence interval Mean	Equivalence interval Lower	Equivalence interval Upper	Reject?	Smallest equivalence interval (±%)	Slope Estimate	Slope 95% C.I. Lower	Slope 95% C.I. Upper	Reject?	Smallest equivalence interval (±%)	No. of obs.
By species and biogeoclimatic zone[a]														
Coast														
Ba, CWH	23.95	23.25	24.62	25.39	22.85	27.93	Y		0.831	0.611	1.072	N	40	203
Cw, CWH	20.92	20.29	21.57	21.36	19.22	23.49	Y		0.551	0.371	0.738	N	65	245
Fd, CDF	29.24	28.30	30.14	30.77	27.69	33.85	Y		-0.544	-0.903	-0.160	N	195	97
Fd, CWH	29.84	29.44	30.22	32.31	29.08	35.54	Y		0.663	0.556	0.778	N	45	884
Hw, CWH	28.02	27.69	28.33	25.53	22.98	28.09	N	15	0.575	0.458	0.697	N	55	1289
Ss, CWH	32.04	31.34	32.74	28.69	25.82	31.56	N	15	0.384	0.047	0.751	N	100	235
Interior														
At, BWBS	18.91	18.50	19.32	17.53	15.78	19.28	N	15	0.351	0.071	0.644	N	95	266
At, ICH	18.61	17.56	19.77	20.55	18.49	22.60	N	15	1.205	0.552	2.234	N	125	35
At, SBS	19.07	18.59	19.56	19.25	17.33	21.18	Y		0.520	0.250	0.795	N	80	215
Bl, ESSF	12.37	11.99	12.75	15.42	13.88	16.96	N	25	0.393	0.231	0.566	N	80	428
Bl, ICH	17.25	16.31	18.15	18.62	16.76	20.48	N	15	0.200	-0.267	0.711	N	130	76
Bl, MS	16.62	15.50	17.77	17.43	15.69	19.17	N	15	-0.761	-1.674	0.124	N	270	68
Bl, SBS	15.30	14.84	15.76	18.47	16.62	20.31	N	20	0.167	-0.010	0.355	N	105	280
Cw, ICH	17.77	17.14	18.39	17.45	15.71	19.20	Y		0.453	0.009	0.890	N	105	175
Ep, ICH	19.19	18.41	19.94	19.52	17.56	21.47	Y		1.026	0.108	1.959	N	100	44
Ep, SBS	18.62	17.90	19.34	18.93	17.03	20.82	Y		0.563	-0.669	1.834	N	175	49
Fd, ICH	21.12	20.72	21.53	23.60	21.24	25.96	N	15	0.210	0.023	0.412	N	100	404
Fd, IDF	16.41	16.16	16.65	18.23	16.41	20.06	N	15	0.920	0.796	1.044	N	25	908
Fd, MS	16.20	15.39	17.01	18.58	16.72	20.44	N	20	0.721	0.122	1.427	N	90	79
Fd, SBS	20.07	19.51	20.63	20.82	18.73	22.90	Y		0.455	0.068	0.937	N	95	143
Hw, ICH	18.00	17.53	18.46	19.08	17.17	20.99	Y		0.534	0.206	0.871	N	80	294
Lw, ICH	22.17	21.53	22.77	23.90	21.51	26.29	Y		0.519	-0.045	1.233	N	110	94
Lw, IDF	19.79	18.94	20.64	20.16	18.14	22.17	Y		0.319	-0.069	0.875	N	110	60
Lw, MS	19.73	18.51	20.88	19.59	17.63	21.55	Y		0.697	0.060	1.543	N	95	35
Pl, BWBS	16.84	16.33	17.36	17.48	15.73	19.23	Y		0.056	-0.245	0.321	N	130	135
Pl, ESSF	15.15	14.72	15.57	16.47	14.83	18.12	N	15	0.549	0.334	0.776	N	70	253
Pl, ICH	20.03	19.70	20.36	21.97	19.78	24.17	N	15	0.720	0.493	0.944	N	55	278
Pl, IDF	15.53	15.17	15.88	18.74	16.87	20.61	N	20	0.844	0.682	1.030	N	35	317

Stratum	Estimate	Intercept							Slope					No. of obs.
		95% C.I.		Equivalence interval			Reject?	Smallest equivalence interval (±%)	Estimate	95% C.I.		Reject?	Smallest equivalence interval (±%)	
		Lower	Upper	Mean	Lower	Upper				Lower	Upper			
Pl, MS	16.70	16.39	17.03	18.39	16.55	20.23	N	15	1.175	0.996	1.358	N	40	428
Pl, SBPS	13.86	13.50	14.22	16.41	14.77	18.05	N	20	0.950	0.828	1.068	N	20	288
Pl, SBS	19.06	18.85	19.27	19.84	17.85	21.82	Y		0.858	0.729	0.989	N	30	731
Py, PP	12.52	11.55	13.59	15.67	14.10	17.24	N	30	3.671	1.126	6.312	N	535	28
Sb, BWBS	10.85	10.15	11.58	10.25	9.23	11.28	N	15	0.630	-0.141	1.431	N	120	70
Se, ESSF	16.07	15.43	16.76	16.14	14.53	17.75	Y		0.240	0.001	0.472	N	100	157
Sw, BWBS	18.62	18.01	19.23	19.08	17.17	20.99	Y		0.364	0.071	0.678	N	95	164
Sx, ICH	21.39	20.81	21.96	22.29	20.06	24.52	Y		0.270	-0.036	0.663	N	105	127
Sx, IDF	16 59	15.32	17.92	18.87	16.98	20.76	N	20	1.605	-0.304	3.057	N	205	28
Sx, MS	17.61	16.73	18.62	18.17	16.35	19.98	Y		0.515	-0.039	0.927	N	105	46
Sx, SBS	19.20	18.90	19.49	20.08	18.07	22.09	Y		0.459	0.248	0.677	N	75	594

a See Table 1 for an explanation of the codes for species and biogeoclimatic zones; model codes are B – biophysical model, S – SIBEC model.

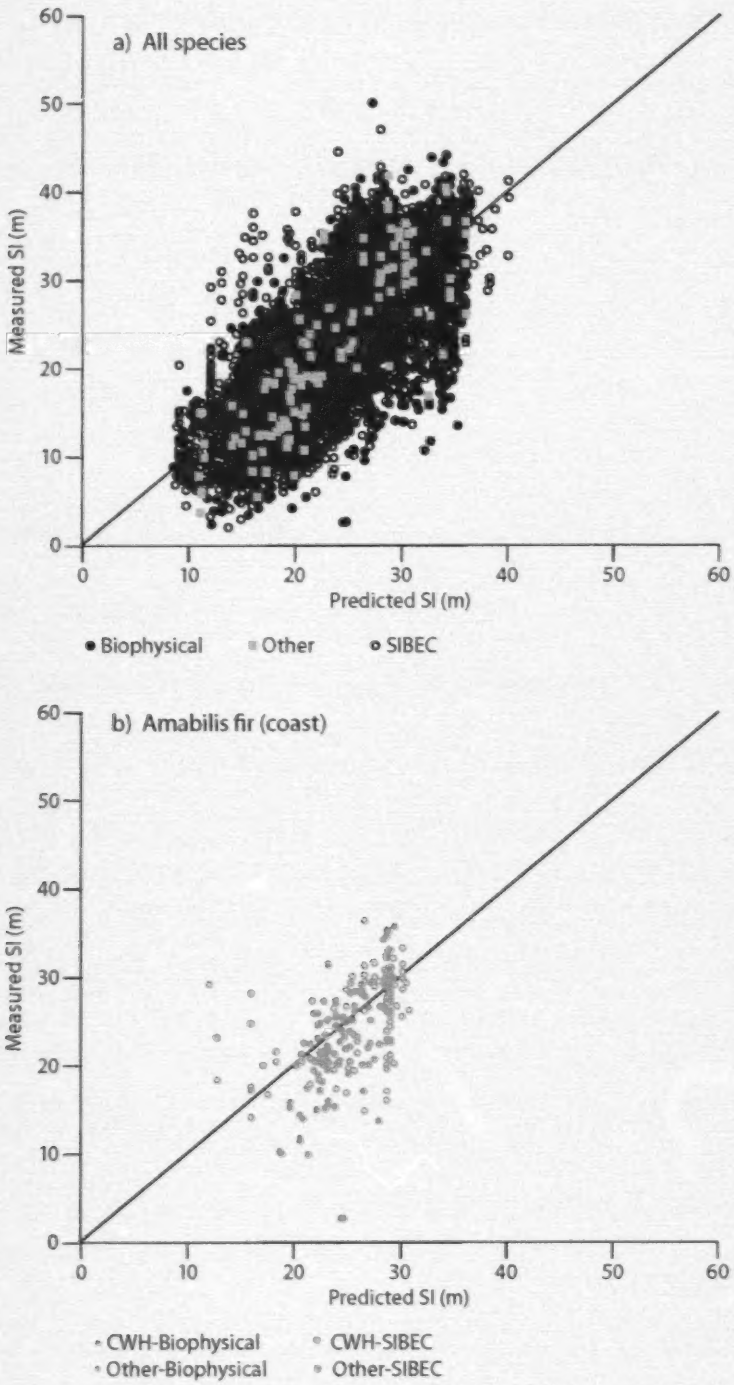
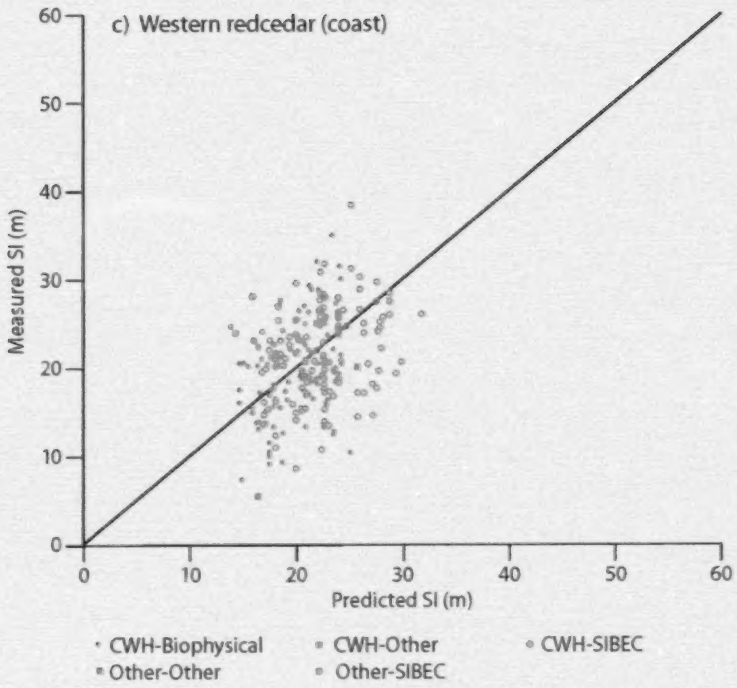


FIGURE 1 *Plots of ground-based site index (measured SI) versus layer site index (predicted SI) by species. Model type (biophysical, SIBEC, or other [i.e., a mixture of the two]) and biogeoclimatic zone are denoted by the symbol type and colour, respectively (see legend; see Table 1 for an explanation of the codes for biogeoclimatic zones).*



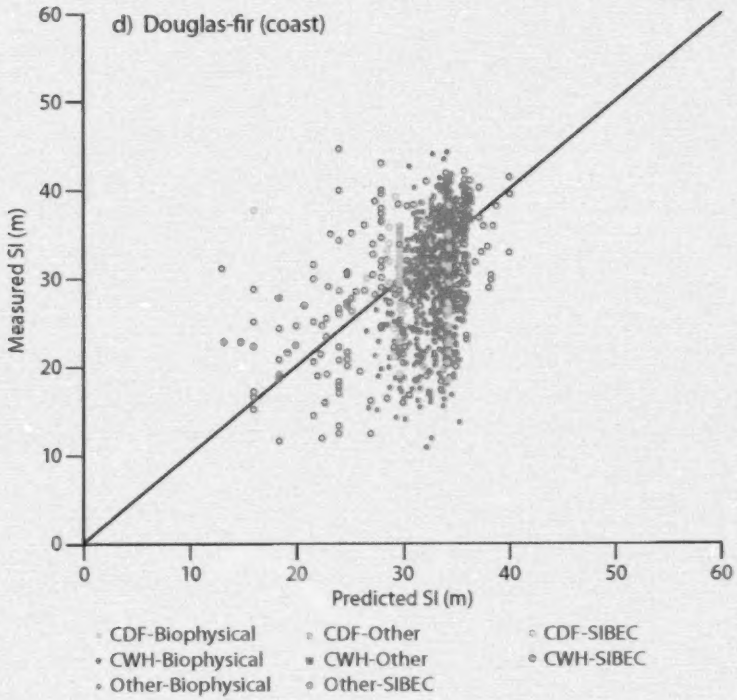
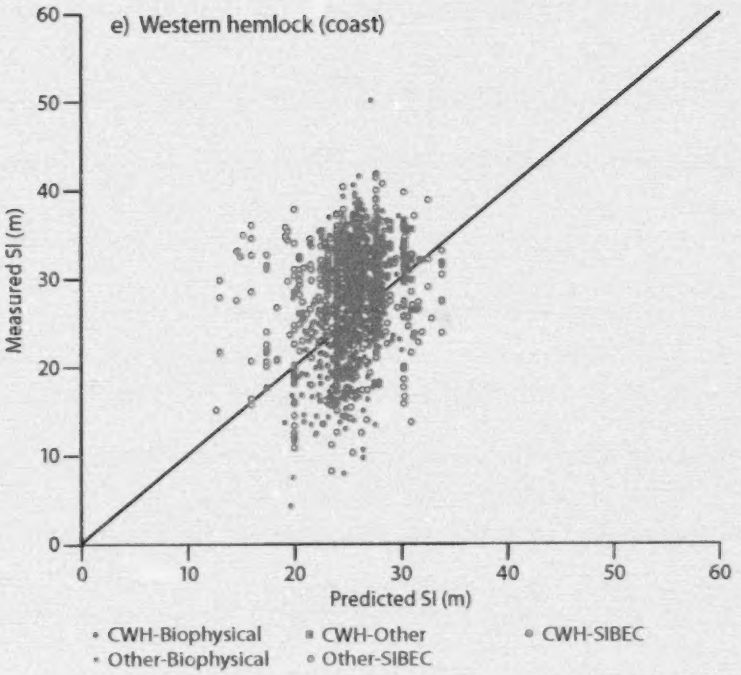


FIGURE 1 *Continued*



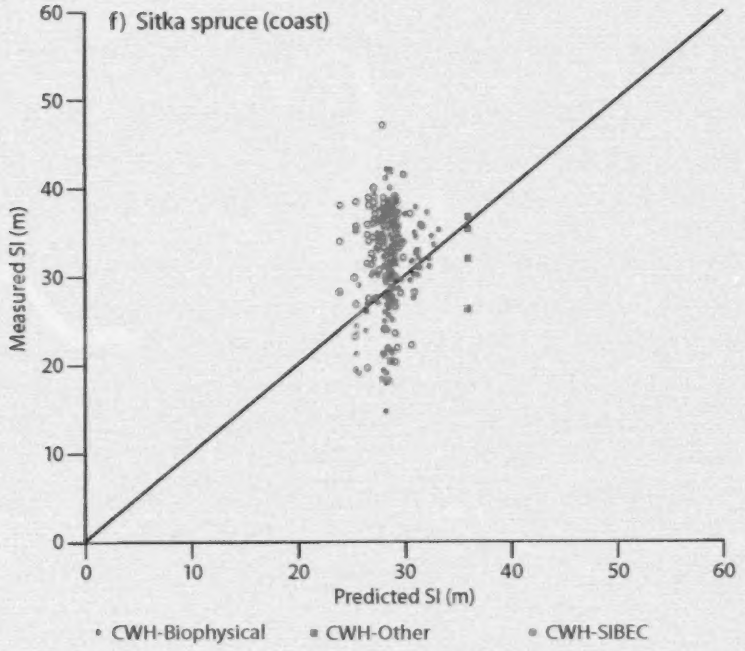
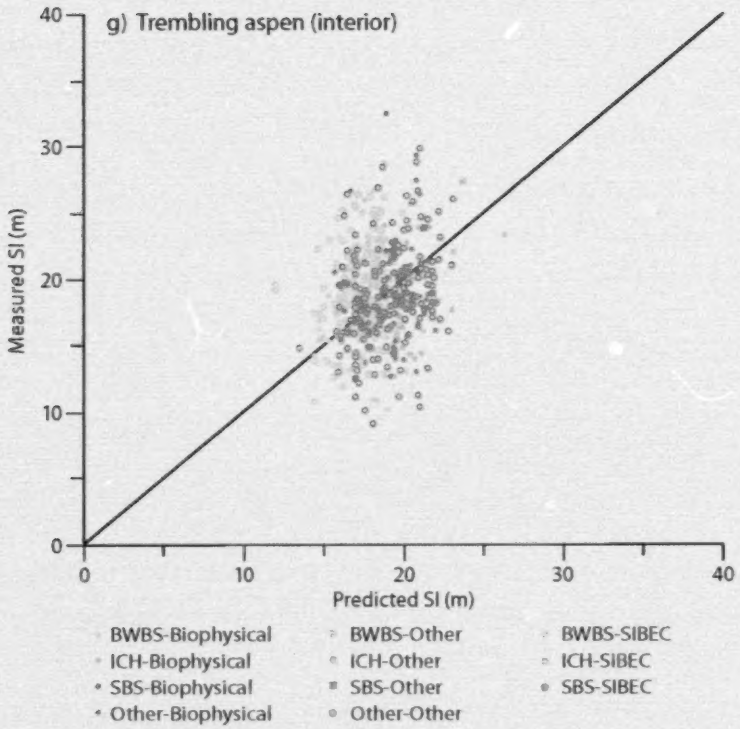


FIGURE 1 *Continued*



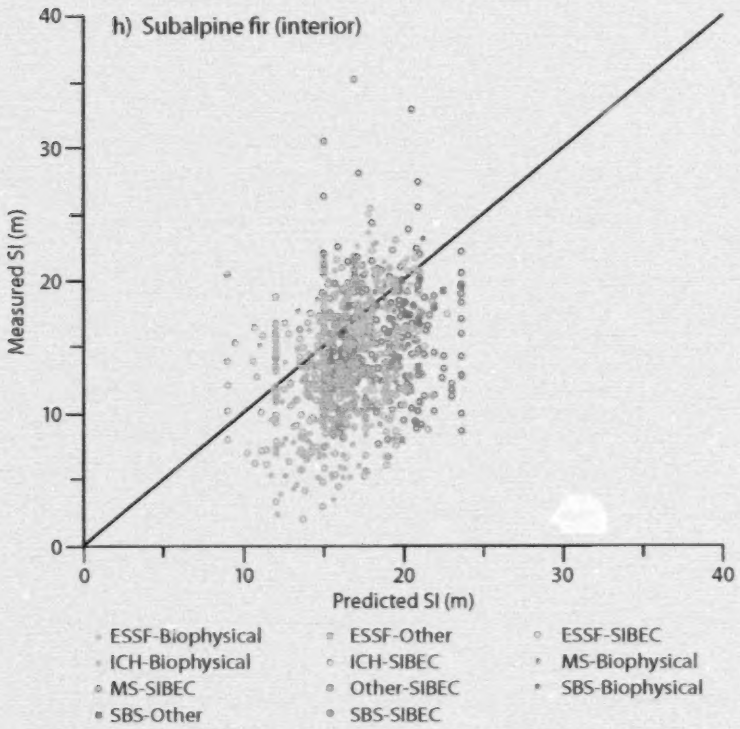
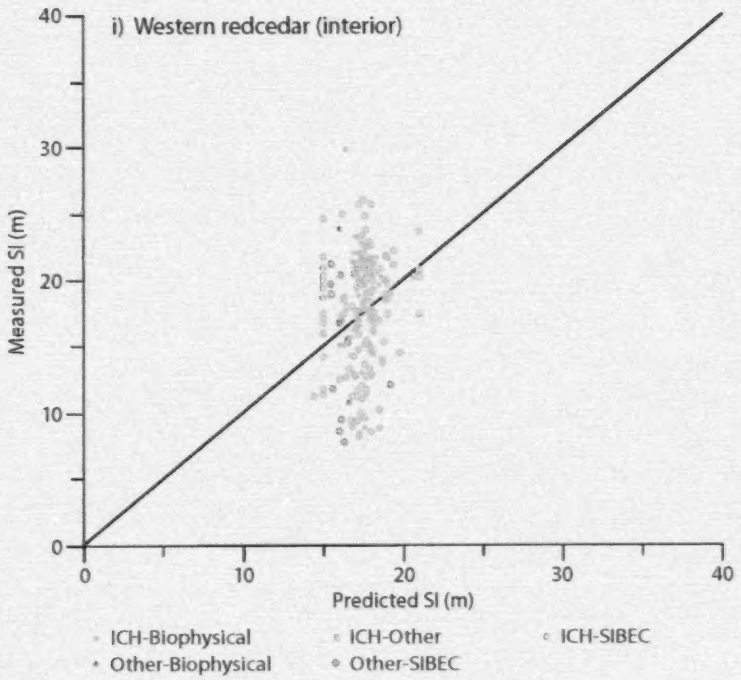


FIGURE 1 *Continued*



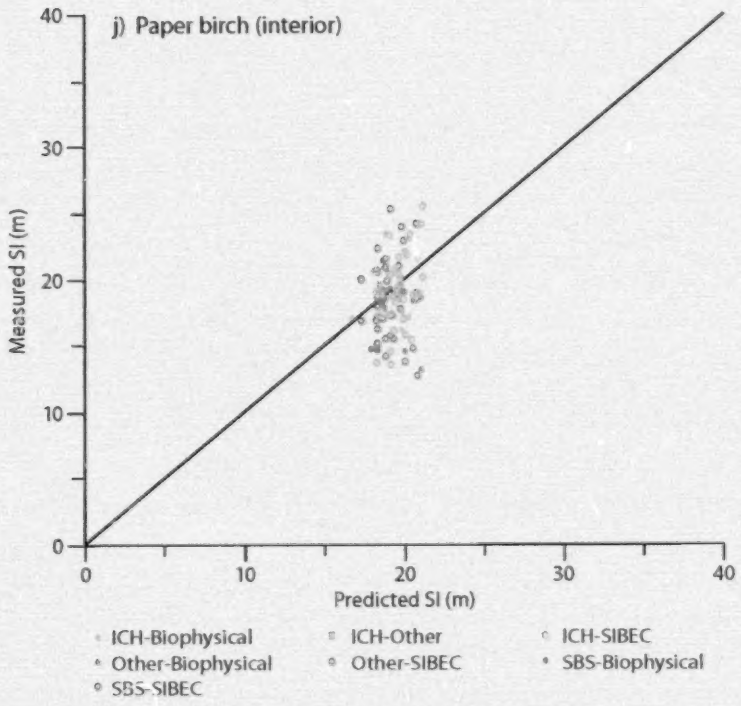
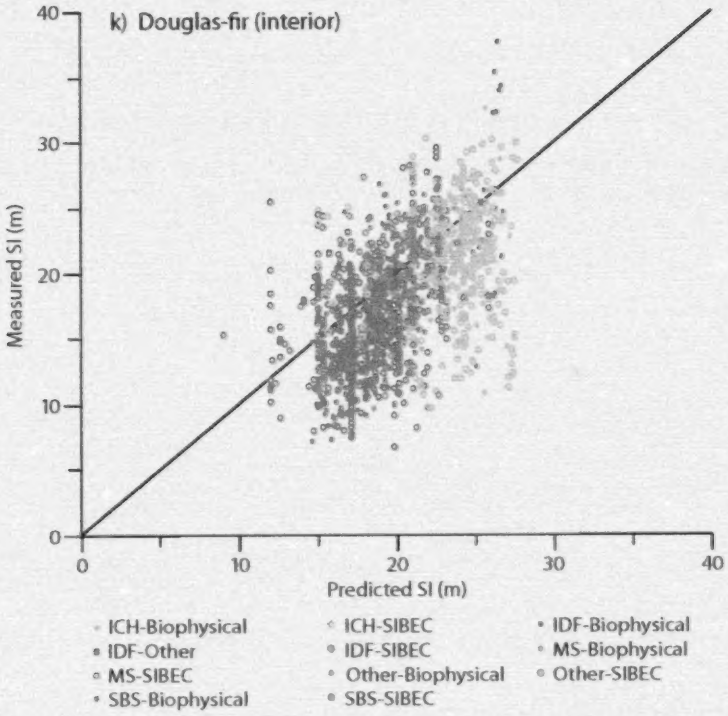


FIGURE 1 *Continued*



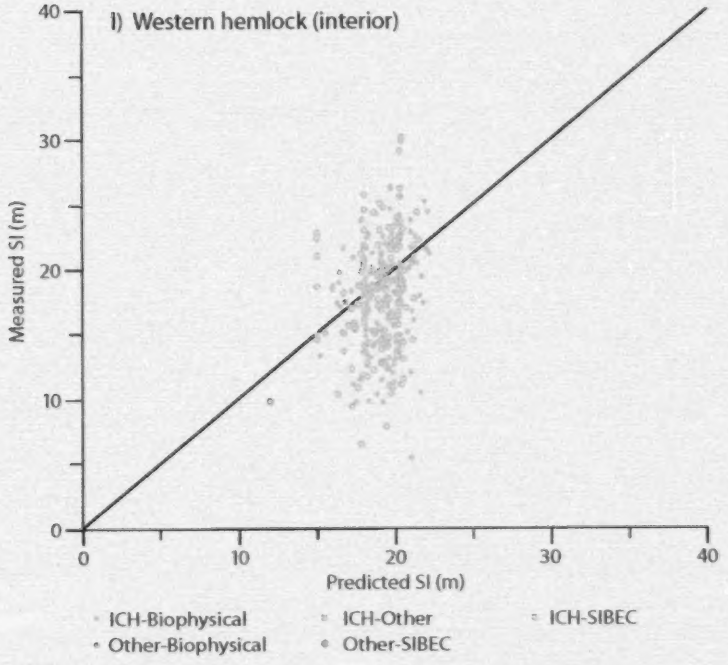
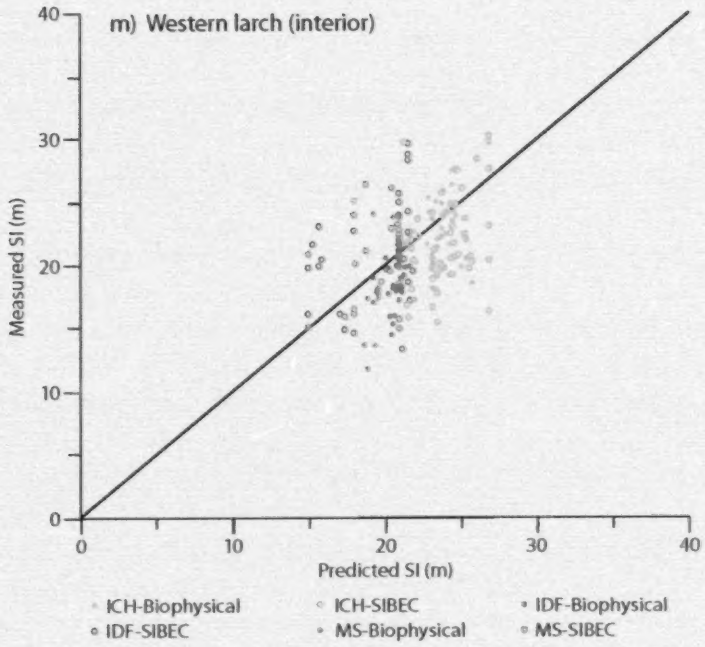


FIGURE 1 *Continued*



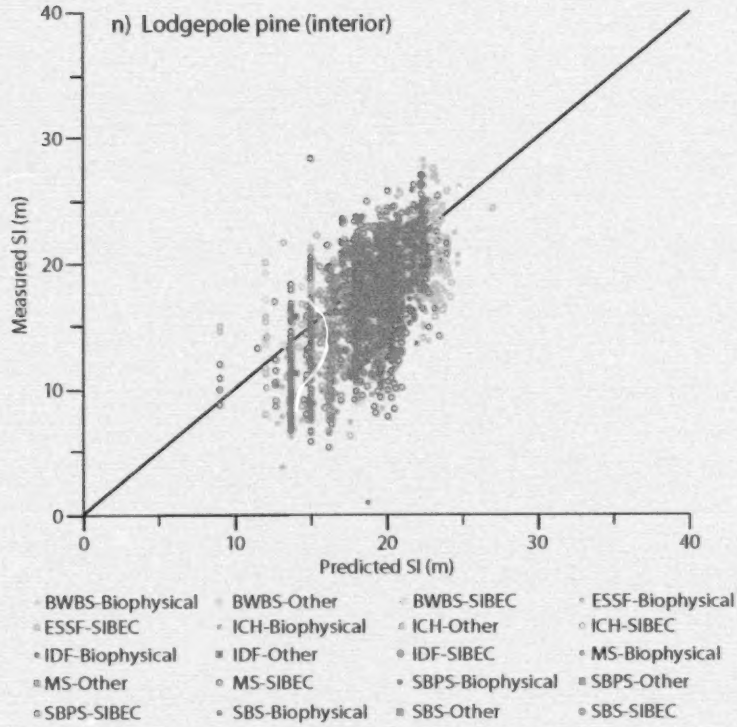
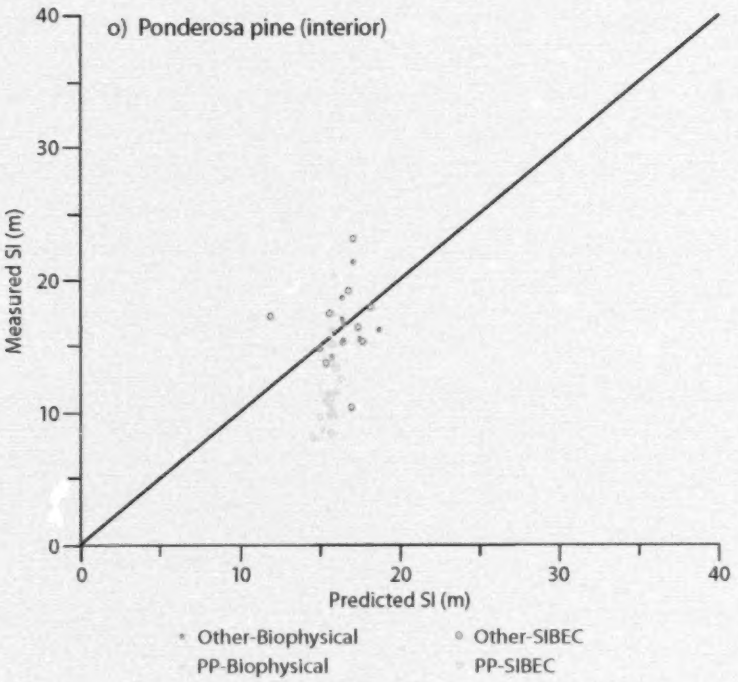


FIGURE 1 *Continued*



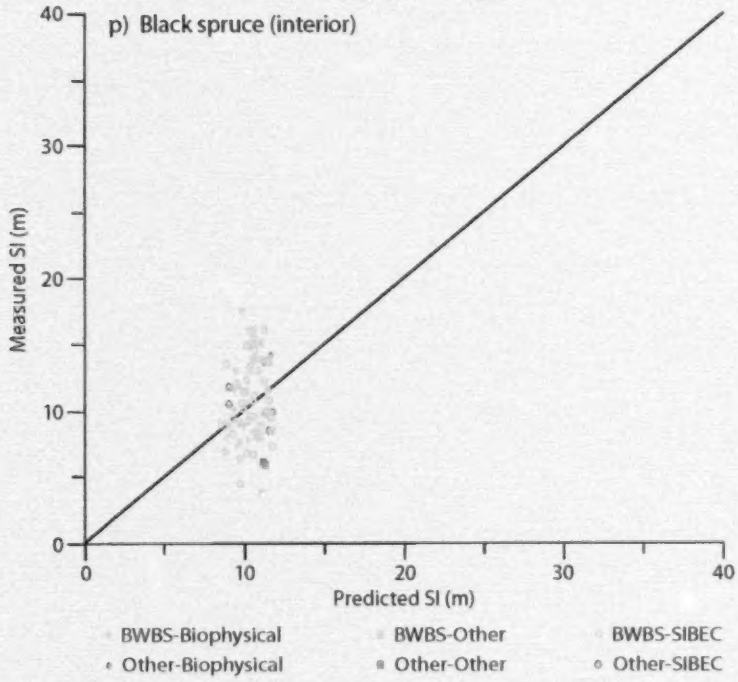
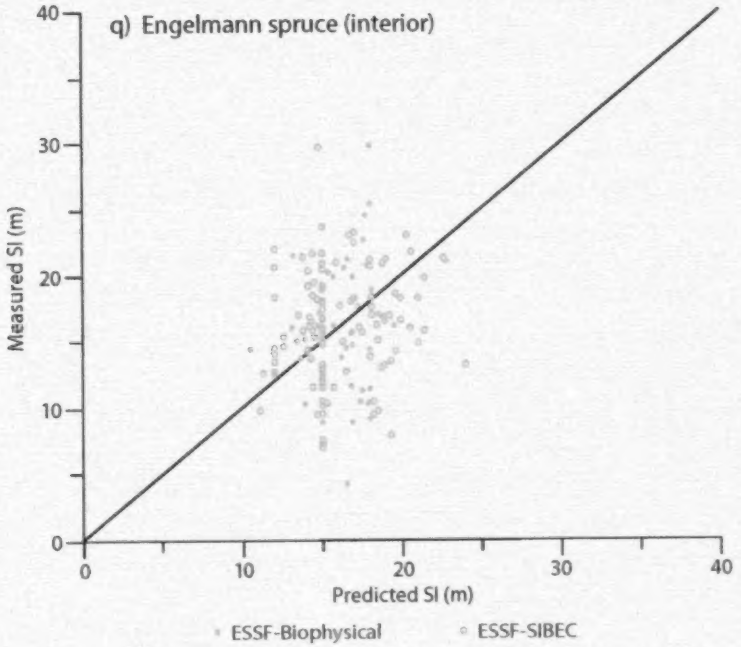


FIGURE 1 *Continued*



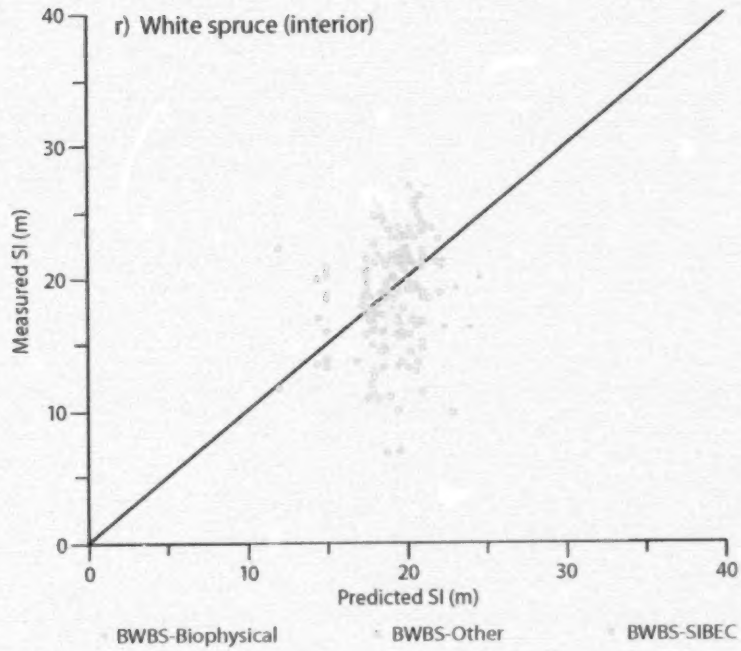


FIGURE 1 *Continued*

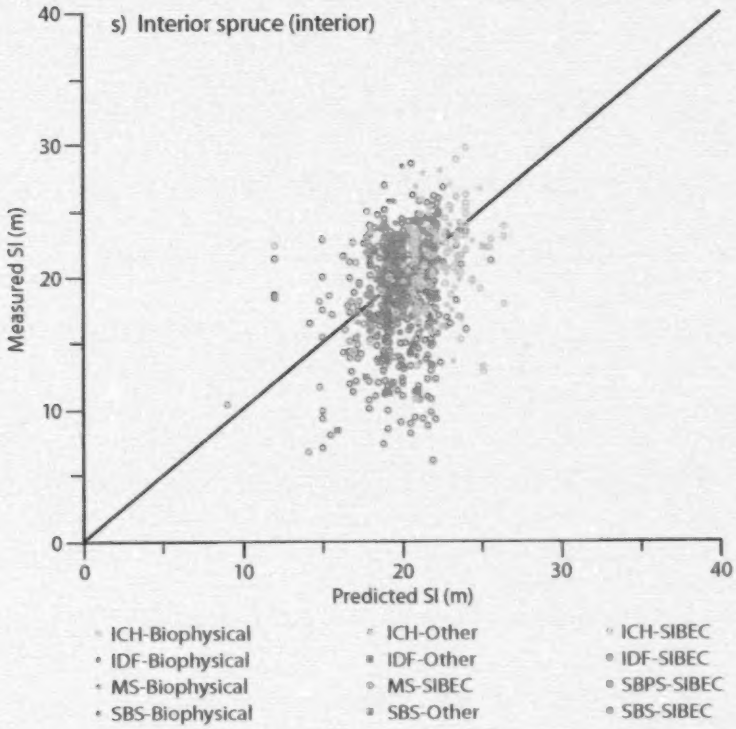


FIGURE 1 *Concluded*

As a whole and by model type, the layer passes the validation. When the analysis is done by stratum, the results are mixed. With a few exceptions, most notably interior Douglas-fir and subalpine fir, and coastal western hemlock, the layer is unbiased for the species strata with a ±10% equivalence interval. The site indexes for all of the species strata are unbiased at a ±20% equivalence interval. The estimated slopes of the relationship between the measured site index and the layer site index, however, vary considerably from 1 for most of the strata. When divided into the species and model type strata and into the species and biogeoclimatic zone strata, many strata are not validated.

For strata where the null hypothesis is not rejected, it is not correct to conclude that the layer is invalidated. It only means that currently there is not enough information to validate the layer.

The choice for the width of the equivalence interval is somewhat arbitrary. Here, it was chosen to be ±10%. This means, for example, that the 95% confidence interval for the mean site index of the observations must be within 10% of the mean predicted site index. Increasing the width of the equivalence interval can lead to the conclusion that the site index layer is validated when the site indexes are in fact substantially biased.

The strata with smaller sample sizes tend to have lower rejection rates for the null hypothesis for the slope and the intercept with a ±10% equivalence interval. All else being equal, a larger sample size will decrease the width of the confidence interval for the intercept and the slope, which in turn will require a smaller equivalence interval in order to reject the null hypothesis (Robinson et al. 2005). However, regardless of how small the confidence interval is, if the parameter of interest is outside of the equivalence interval, then the null hypothesis of dissimilarity will not be rejected.

There are many sources of error both in the layer site indexes (predictions) and the ground-based site indexes (observations) that may lead to the failure to reject the null hypothesis. The following is a non-exhaustive accounting of these errors:

1. The ground-based site indexes are estimated from measured height and age, so there is model prediction error in these site indexes and some measurement error.
2. The pairs of data consist of a site index for a 1-ha square cell from the layer and a measured site index from a plot with a size that is considerably smaller than the 1-ha cell. Spatial heterogeneity of site index within the cell may lead to differences between the two site indexes.
3. The SIBEC and biophysical models are analyzed with both plot data and stem analysis data. The site indexes from the plot data are predicted from the height and age of a site tree, and hence have model prediction error, which can be large, and measurement error. Furthermore, the variance of the errors in the model predictions depends on the age of the site tree (see Nigh 1998 for an analysis of this issue). The site indexes from stem-analyzed trees are measured directly and do not have model prediction error, but only a small proportion of the trees in the SIBEC and biophysical data sets are stem-analyzed trees.

4. A PEM or TEM model is used to apply the SIBEC site indexes. A TEM or PEM needs to attain only a 65% accuracy level (Meidinger 2003) to be used in the layer. This will lead to errors in site indexes for the layer since an error in the PEM or TEM will lead to the incorrect site index assignment to the layer cell.
5. Both the biophysical model and the SIBEC model have large variances for individual predictions. Therefore, using the layer to estimate site index for small-scale applications is not reasonable. These large variances represent variability that is not captured by the model and can be due to known factors (e.g., soil moisture/nutrient regime for the biophysical model, elevation for the SIBEC model) or unknown factors (see Nigh 2010 for an analysis of the variability within a site series).
6. The SIBEC model predicts the identical site index for each species/site series combination, so no within-site series variability is captured by this model. The vertically aligned dots in the graphs (e.g., see the graph for lodgepole pine) are likely SIBEC site index estimates from one site series.
7. The selection of site trees from many of the plots was done as an office exercise. It is likely that some of these trees will not be site trees due to suppression, damage, or many other factors, especially in older trees. This will usually create a downwards bias in the estimated site index.

CONCLUSION

Equivalence testing was used to validate the site productivity layer, which as a whole and by model type passes the validation. However, when the layer is divided into smaller strata, many of the individual strata do not pass the validation test. In most cases, the individual strata are validated in terms of overall bias but do not pass the validation test for correspondence between the predicted and actual site indexes.

Barker, L.E., E.T. Luman, M.M. McCauley, and S.Y. Chu. 2002. Assessing equivalence: an alternative to the use of difference tests for measuring disparities in vaccination coverage. Am. J. Epidemiol. 156:1056–1061.

Barker, L., H. Rolka, D. Rolka, and C. Brown. 2001. Equivalence testing for binomial random variables: which test to use? Am. Stat. 55:279–287.

Blair, R.C. and S.R. Cole. 2002. Two-sided equivalence testing of the difference between two means. J. Mod. Appl. Stat. Methods 1:139–142.

British Columbia Ministry of Environment. 2014. Terrestrial & predictive ecosystem mapping home. Victoria, B.C. www.env.gov.bc.ca/ecology/tem/ (Accessed Sept. 23, 2014).

British Columbia Ministry of Forests. 1994. A review of operational growth and yield monitoring in British Columbia. Victoria, B.C. www.for.gov.bc.ca/hts/vri/technical/growthyield/gymonitoring/opergymonitoring1994.pdf (Accessed Sept. 24, 2014).

British Columbia Ministry of Forests and Range. 2001. SiteTools, version 3.3. Victoria, B.C. www.for.gov.bc.ca/hre/sitetool/ (Accessed Oct. 9, 2014).

_____. 2011. Site index estimates by site series (SIBEC) - second approximation. Victoria, B.C. www.for.gov.bc.ca/hre/sibec/ (Accessed Sept. 23, 2014).

_____. 2014. Biogeoclimatic ecosystem classification program. Res. Br., Victoria, B.C. www.for.gov.bc.ca/hre/becweb/index.html (Accessed Dec. 11, 2014).

British Columbia Ministry of Forests, Lands and Natural Resource Operations. 2014a. Provincial site productivity layer. Victoria, B.C. Updated July 22, 2014. www.for.gov.bc.ca/hts/siteprod/provlayer.html (Accessed Sept. 22, 2014).

_____. 2014b. Vegetation resources inventory monitoring program national forest inventory and monitoring. Victoria, B.C. Updated June 26, 2014. www.for.gov.bc.ca/hts/vri/monitoring/monitoring.html (Accessed Sept. 23, 2014).

_____. 2014c. Vegetation Resources Inventory permanent sample plots. Victoria, B.C. Updated Mar. 6, 2014. www.for.gov.bc.ca/hts/vri/psps/psp.html (Accessed Sept. 23, 2014).

_____. 2014d. Vegetation Resources Inventory. Victoria, B.C. Updated June 19, 2014. www.for.gov.bc.ca/hts/vri/ (Accessed Sept. 23, 2014).

Centre for Forest Conservation Genetics. 2014. ClimateBC/WNA. http://cfcg.forestry.ubc.ca/projects/climate-data/climatebcwna/#ClimateWNA (Accessed Nov. 4, 2014).

CloverPoint. 2014. FLNR provincial site productivity layer PEM/TEM-SIBEC and biophysical analysis. Ver. 4.3. www.for.gov.bc.ca/hts/siteprod/download/FLNR_provincial_site_product_layer_pem-tem-sibec_biophysical_analysisv43.pdf (Accessed Sept. 24, 2014).

Hectares BC. 2014. Hectares BC. www.hectaresbc.org/app/habc/HaBC.html (Accessed Sept. 22, 2014).

Meidinger, D. 2003. Ecosystem mapping accuracy and timber supply applications. B.C. Min. For., Res. Br., Victoria, B.C. www.for.gov.bc.ca/hre/becweb/Downloads/Downloads_PEM/PEMaccuracystatement2003.pdf (Accessed Apr. 15, 2014).

Meidinger, D. and J. Pojar. 1991. Ecosystems of British Columbia. B.C. Min. For., Res. Br., Victoria, B.C. Spec. Rep. 6.

Nigh, G.D. 1998. A system for estimating height and site index of western hemlock in the interior of British Columbia. For. Chron. 74:588–596.

_____. 2010. A closer look at site index – biogeoclimatic site series correlations: Douglas-fir in the Coastal Western Hemlock zone, xm2 variant, 01 site series. For. Chron. 86:477–483.

_____. 2012. A biophysical model for estimating site index for the major commercial tree species in British Columbia. Prov. B.C., Victoria, B.C. Tech. Rep. 073.

Parkhurst, D.F. 2001. Statistical significance tests: equivalence and reverse tests should reduce misinterpretation. BioSci. 51:1051–1057.

R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.

Robinson, A.P., R.A. Duursma, and J.D. Marshall. 2005. A regression-based equivalence test for model validation: shifting the burden of proof. Tree Physiol. 25:903–913.

Robinson, A.P. and R.E. Froese. 2004. Model validation using equivalence tests. Ecol. Modelling 176:349–358.

Rykiel, E.J., Jr. 1996. Testing ecological models: the meaning of validation. Ecol. Modelling 90:229–244.

Wang, T., A. Hamman, D.L. Spittlehouse, and T.Q. Murdoch. 2011. ClimateWNA—high-resolution spatial climate data for western North America. J. Appl. Meteorol. Climatol. 51:16–29. DOI: 10.1175/JAMC-D-11-043.1.

Wellek, S. 2010. Testing statistical hypotheses of equivalence and noninferiority. CRC Press, New York, N.Y.